Content Syndication

philip@karova.com

What is Content Syndication?

- *Content Syndication* makes part or all of a site's content available for use by other services. The syndicated content usually referred to as a feed, can consist of both the direct content and metadata describing that content.

  The feed can contain anything from just headlines and links to stories, excerpts of stories or even the entire content of the site, stripped of it's layout.

- This presentation will be discussing Content Syndication with an emphasis on the RSS and ATOM technologies.

The History of Content Syndication

1995

The very first version of a content syndication format was MCF (Meta Content Framework). A project funded by Apple whose aim was to describe objects, attributes, and the relationships between them.

1997

The project lead for MCF then moved to Netscape where MCF was moved over to an XML-based format. This later became known as RDF (Resource Description Framework).

The History of Content Syndication

RDF became a standard recognised by the W3C (World Wide Web Consortium) who described it as “a general-purpose language for representing information on the World Wide Web”. It is specifically designed for the representation of metadata and the relationships between things. RDF is the basis for the concept known as the “Semantic Web”, the W3C’s vision of a web of information that computers can understand.

The History of Content Syndication

1997

As this was the height of the Microsoft and Netscape wars and Netscape had just released RDF, Microsoft released CDF (Channel Definition Format), which was very similar to RDF in what it provided.

1999

Netscape launch the “My Netscape Network” portal and with it a new format called RSS (RDF Site Summary) which was a subset of RDF, RSS allowed the new portal to display headlines and urls from other sites all within the same page.

The History of Content Syndication

Users could then personalise their Netscape portal page to contain the headlines from any site that interested them and that had an RSS feed available. This was the first large scale web-based RSS aggregator (news reader). It was a one-stop shop for a day's browsing. RSS (version 0.90) was so popular that it didn't stay unique to Netscape for long

While Netscape's portal was being developed, another similar project was being developed called "Radio Userland", this project had developed a new format based on RDF called "ScriptingNews".

The History of Content Syndication

Later versions of ScriptingNews took ideas from RSS 0.90 (RSS now stood for Rich Site Summary) and added features. Netscape saw this and Userland and Netscape worked together to create RSS 0.91.

2000

Work continued on the specs for RSS 0.91 and eventually RSS 0.92 was released which included optional elements like enclosure.

RSS 1.0 was also developed but was deemed to complex compared to the 0.9x versions. It used RDF for it's metadata descriptions.

The History of Content Syndication

2002

RSS 2.0 (Now an acronym for Really Simple Syndication) was released. This and variations of it, is the format used today.

The reason for all the different versions was that each version added new features like author, comments etc.

The History of Content Syndication

2003

ATOM 0.3 was released by the IETF (Internet Engineering Task Force) which was developed in response to the confusing terminology and versioning of the flavors and approaches used in RSS. The main difference was that it used namespaces unlike RSS 2.0, this allowed the use of Modules like Dublin Core (Metadata) as in RSS 1.0 .

2005

ATOM 1.0 released which introduces a new protocol for feed subscription and publication.

What types are used today?

Main syndication types used today
RSS 0.91
RSS 2.0
ATOM 0.3

RSS 2.0 (Really Simple Syndication)

RSS 2.0 as with the 0.9x versions does not use RDF for metadata and as such is a very simple format and very easy to pick up and even easier to parse.

The root element is ‘rss’. The child of ‘rss’ that is permitted is ‘channel’. The required child elements of ‘channel’ are ‘title’, ‘link’ and ‘description’. These contain metadata for the channel. Each channel will have multiple ‘item’s.

RSS 2.0 (Really Simple Syndication)

Each item has it's content inside the description element or inside the content:encoded element (part of the content module). An item can have a 'category', 'author', 'pubdate' etc.

Two great features of RSS are the 'comments' element which points the reader to comments related to the item. The other features is enclosure which attaches another media type to the item much like attachments in emails.

RSS (Really Simple Syndication) - Example Feed RSS 2.0

```
- <item>
    <title>O'Reilly: What Is Web 2.0</title>
  - <link>
      http://www.philroche.net/archives/oreilly-what-is-web-20/
    </link>
  - <comments>
      http://www.philroche.net/archives/oreilly-what-is-web-20/#comments
    </comments>
    <pubDate>Sat, 01 Oct 2005 12:11:41 +0000</pubDate>
    <dc:creator>philroche</dc:creator>
    <category>General</category>
  - <guid>
      http://www.philroche.net/archives/oreilly-what-is-web-20/
    </guid>
  - <description>
      <p><a href="http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html">O&#82
      Exactly, What Is Web 2.0? </p> <br class="clear" />
    </description>
  - <content:encoded>
      <p><a href="http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html">O&#82
      Exactly, What Is Web 2.0? </p> <br class="clear" />
    </content:encoded>
  - <wfw:commentRSS>
      http://www.philroche.net/archives/oreilly-what-is-web-20/feed/
    </wfw:commentRSS>
  </item>
```

ATOM

As you've probably figured Atom is an XML-based content and metadata syndication format and an application-level protocol for publishing and editing web resources.

Atom is very similar to RSS but it the data is marked up in a different way using with the root element being 'feed' instead of 'rss'. Atom provides metadata about the feed using the 'id', 'title', 'updated', 'author' and 'link'. A feed can also belong to a category. Where rss has 'item' for a feed entry, Atom has 'entry'.

ATOM

An entry must have an author, content, link and summary. Separate to the feeds category, each entry can also be categorised using the category element. One nice feature of Atom is the Atom protocol which is designed to be a standard way for blogging-related software to communicate with the server that is hosting the feed.

Atom has it's own namespace. This allows Atom to be extended as with Modules in RSS.

ATOM - Example Feed

```
- <entry>
  - <author>
      <name>philroche</name>
    </author>
    <title type="text/html" mode="escaped">O&#8217;Reilly: What Is Web 2.0</title>
    <link rel="alternate" type="text/html" href="http://www.philroche.net/archives/oreilly-what-is-web-20/" />
  - <id>
      http://www.philroche.net/archives/oreilly-what-is-web-20/
    </id>
    <modified>2005-10-01T12:11:41Z</modified>
    <issued>2005-10-01T12:11:41Z</issued>
    <dc:subject>General</dc:subject>
  - <summary type="text/plain" mode="escaped">
      O&#8217;Reilly: What Is Web 2.0 Exactly, What Is Web 2.0?
    </summary>
  - <content type="text/html" mode="escaped" xml:base="http://www.philroche.net/archives/oreilly-what-is-w
      <p><a href="http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html">O&#82
      Exactly, What Is Web 2.0? </p> <br class="clear" />
    </content>
  </entry>
```

Are there advantages of one type over the other?

RSS	ATOM
Enclosures	Simple
Widely supported	Standards Body
Simple	Smaller payload
Extensible	Extensible
	Includes protocol

How do we use these feeds?

Aggregators:

RSS and Atom feeds automate the transmission of certain types of data through information feeds. However the goal of these feeds is usually to present data from information feeds to human readers.
The term ‘aggregrator’ is used to refer to a desktop or online tool that manages and displays several information feeds.

Most aggregators also have features that allow automatic updating of feeds, at set intervals or according to the feeds update frequency. This means the user can browse offline.

How do we use these feeds?

Aggregators like Aggreg8 and Habari-Xenu are integrated in the web browser, Others like GreatNews or RSSbandit are standalone applications. There are also online aggregators like Bloglines. The email client Thunderbird also includes RSS and Atom support. Recently browsers themselves have started including aggregators :- Firefox Live Bookmarks, Safari RSS, Opera and IE7.

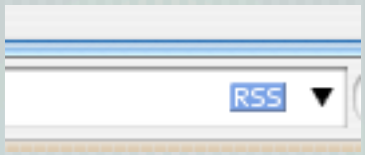
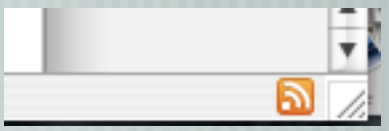
I personally prefer the aggregators that are integrated into the browser as it means one less application to run. It also means that the aggreator can avail of a browsers AutoDiscovery features with a nicer user experience than the included readers.

What is AutoDiscovery?

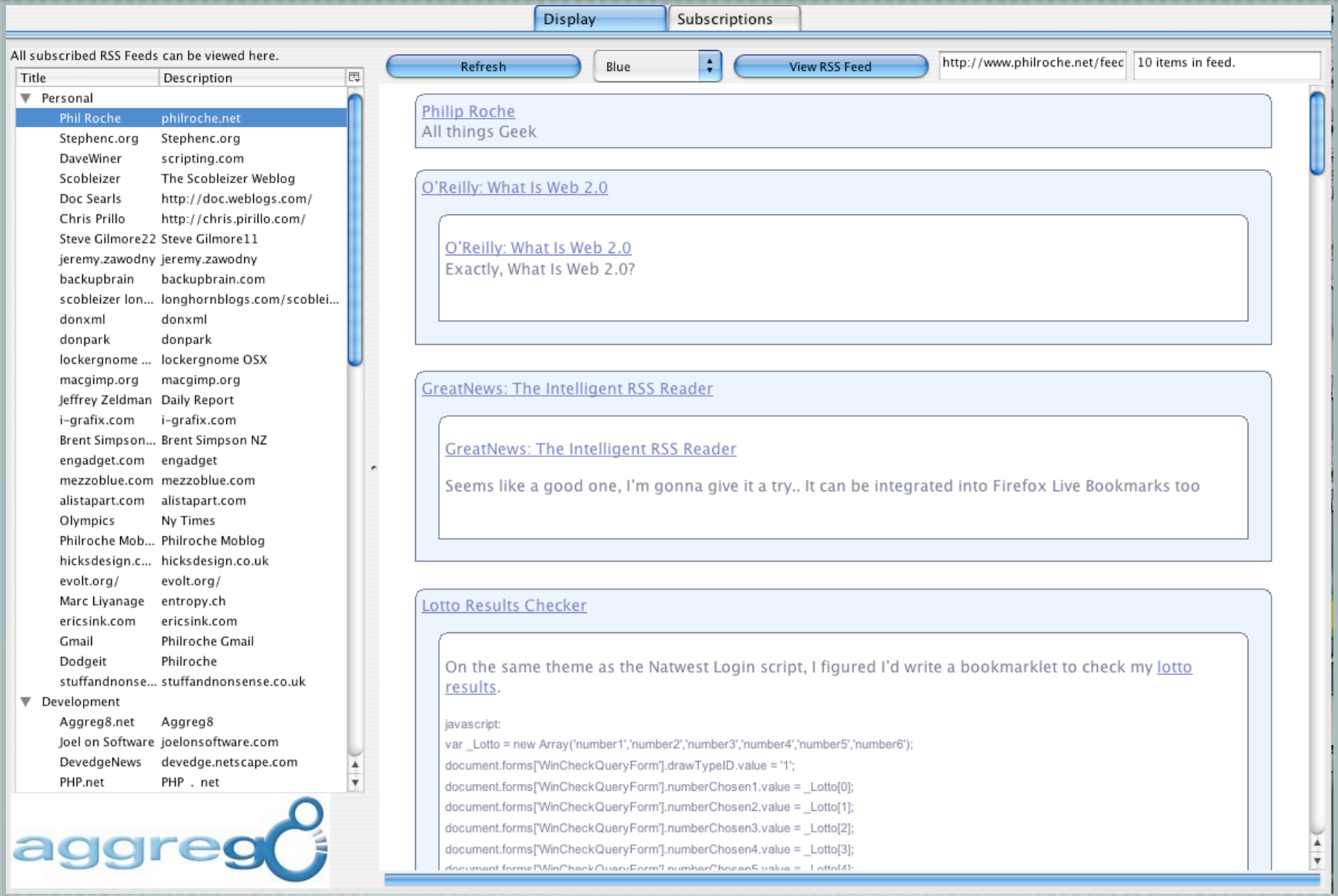
AutoDiscovery is a feature that informs the browser or aggreagtor that a feed is available for the site currently being viewed. Simply include

```
<link rel="alternate" type="application/rss+xml" title="RSS 2.0" href="http://www.philroche.net/feed/" />
<link rel="alternate" type="text/xml" title="RSS .92" href="http://www.philroche.net/feed/rss/" />
<link rel="alternate" type="application/atom+xml" title="Atom 0.3" href="http://www.philroche.net/feed/atom/" />
```

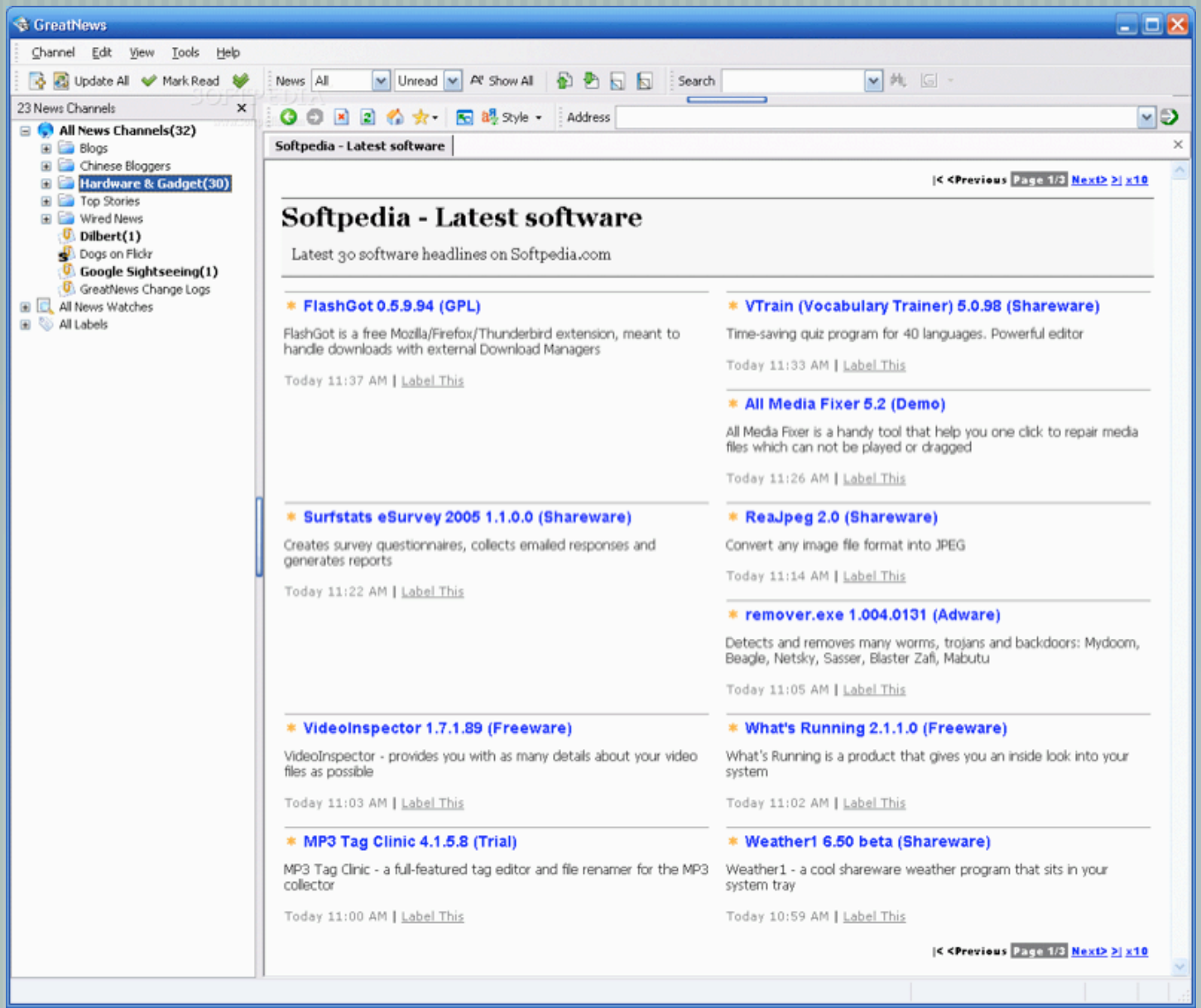
and the browsers/aggregators will recognise that a feed is available.

ScreenShots	Safari	Opera	Firefox	IE7
				

How do we use these feeds? - Aggreg8



How do we use these feeds? - GreatNews



Overview

Up to now we have discussed news feeds but there is a new trend of including other media types in syndication feeds. This can be images, video but most prominent is audio content being included. This has been given the term ‘podcasting’ after the very popular iPod.

Video has also started to appear in feeds which is now called videoblogging or vodcasting.

Enclosures

Enclosures has been a feature in RSS since version 0.92

```
<enclosure url="http://cyber.law.harvard.edu/blogs/audio/lydon/west.emerson.1.mp3" length="6912627" type="audio/mpeg"/>
```

RSS 1.0 can have enclosure but through the use of an enclosure module. Atom can also use this module but a simpler way is to use the Atom link element.

```
<link rel="enclosure" type="audio/mpeg" title="MP3" href="http://www.philroche.net/myaudiofile.mp3" length="1234"/>
```

In the future it is hoped that true enclosure where the binary date is serialised in the feed content.

Media RSS

Media RSS is a new RSS module developed by Yahoo that supplements the enclosure capabilities of RSS 2.0. Enclosures are already being used to syndicate audio and images. Media RSS extends enclosures to handle other media types, such as short films or TV, in addition to providing additional metadata with the media.

It was developed to make it easier for people to find and view rich media content. Media RSS is being integrated into Yahoo's new Video search engine.

Podcasting

Think how a desktop aggregator works. You subscribe to a set of feeds, and then can easily view the new stuff from all of the feeds together, or each feed separately.

Podcasting works the same way, with one exception. Instead of reading the new content on a computer screen, you listen to the new content on an iPod or iPod-like device. Podcasts can be syndicated through RSS or Atom and can come in any audio format but MP3 is the most popular.

Podcasting

In the last few months podcasting has become extremely popular. It is not surprising considering the variety of different topics that are covered:

* Music
* Talk Shows
* Training
* Story
* Comedy
* Interviews

Other non-news syndication types

There are other xml-based syndication formats that, handle different types of data. These are not directly related to RSS and Atom but an aggregator could be used to parse the feeds. Two of these are:

- iCAL
  - Used for calendar syndication and calendar synchronisation.
- XBEL (XML Bookmark Exchange Language)
  - Bookmark Syndication and synchronisation.
- OPML (Outline Processor Markup Language)
  - Used for storing a persons feed subscriptions (blogroll)

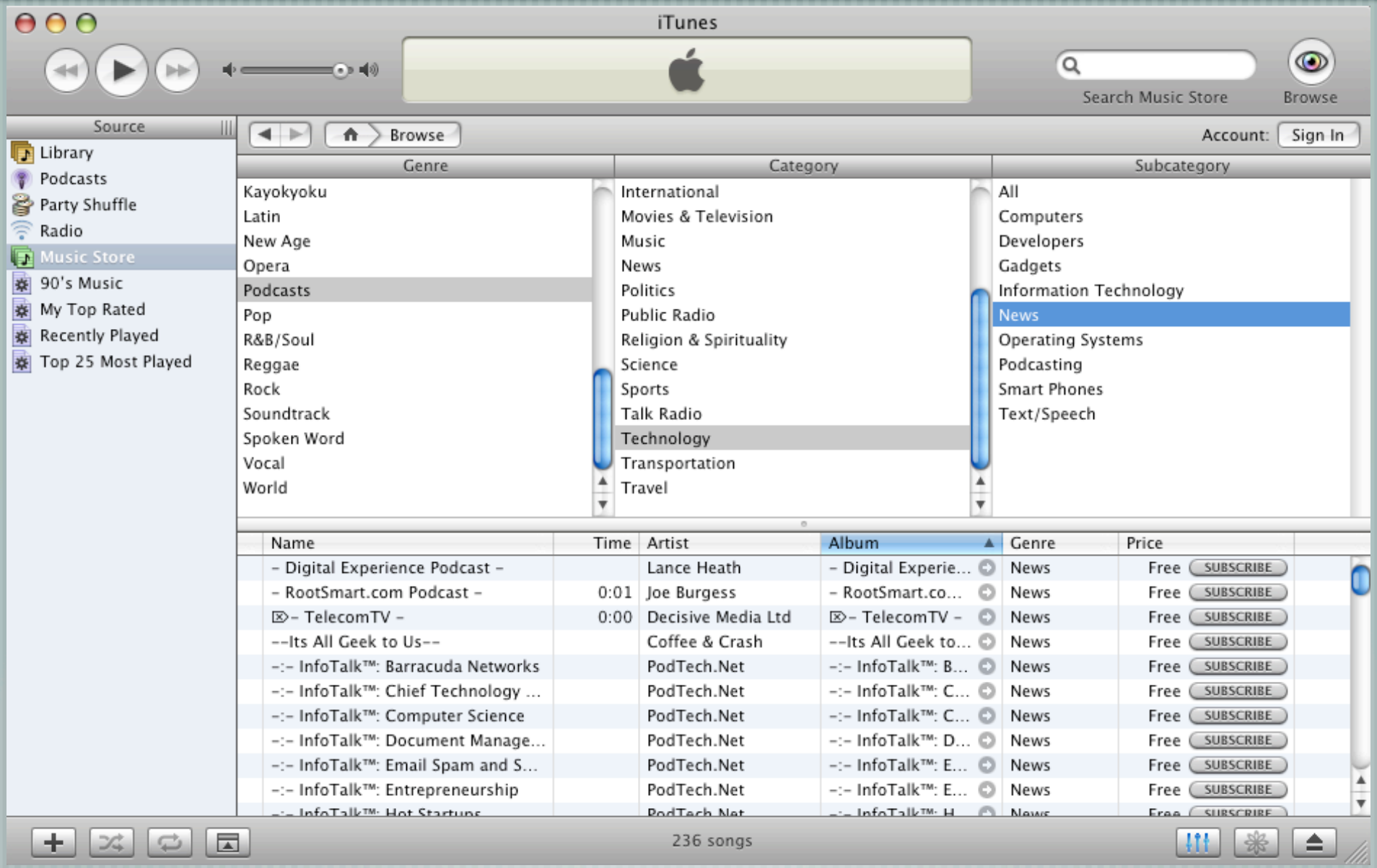
How do we use these feeds?

The feed types and formats discussed in this section can be aggregated and managed by some of the aggregators discussed earlier.

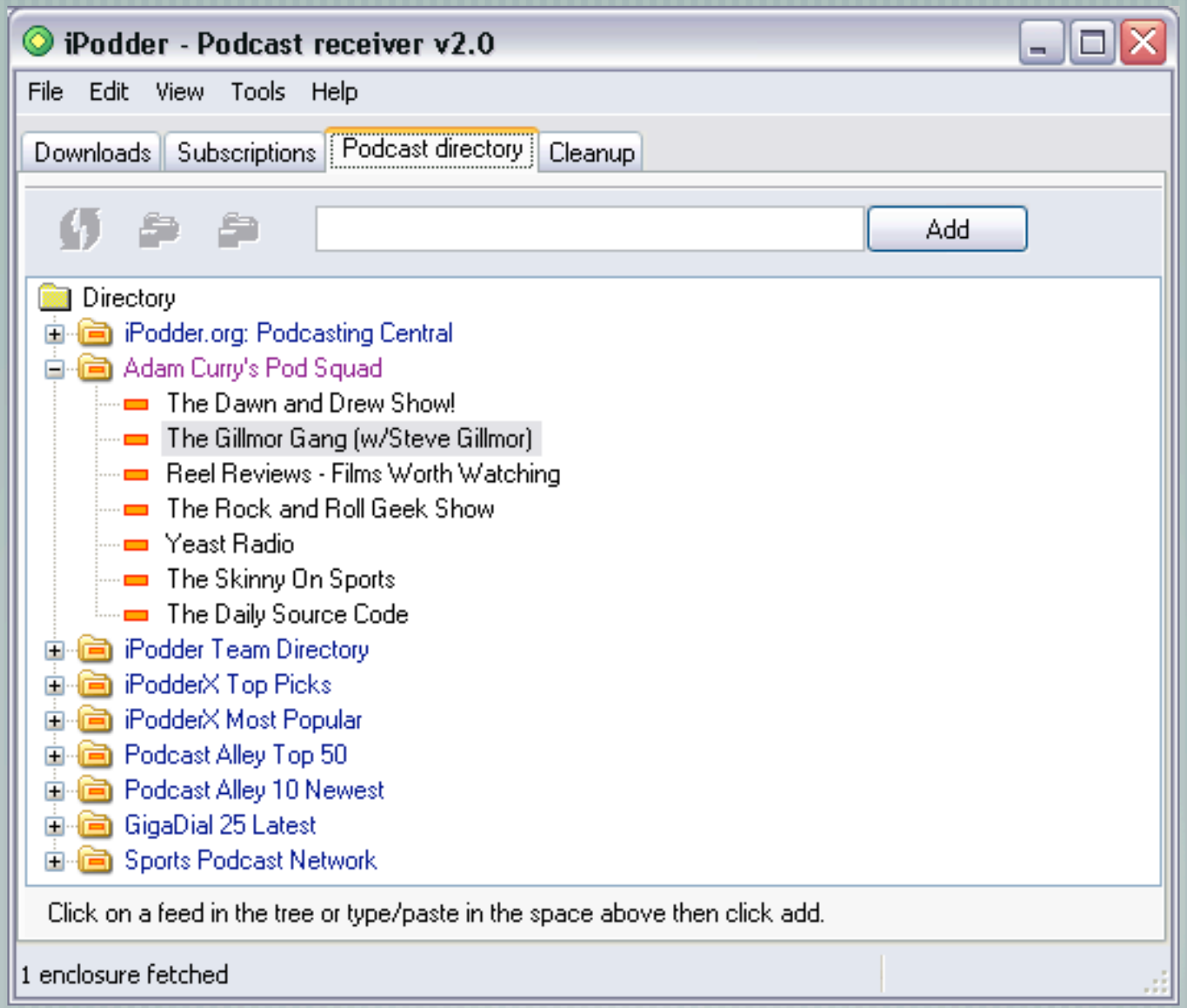
With podcasting however, audio is required. The best podcast managing software available is iTunes which has a complete podcast directory where you can choose your podcasts. These can then be synced to your ipod for later listening or saved for later.

iPodder and iPodderx are also worth metioning as they are cross platform and are not tied to the iPod.

How do we use these feeds?



How do we use these feeds?



Search

Where do you find feeds to subscribe to?

There are numerous search engines that only search blog content, google.com/blogsearch, feedfinder.feedster.com. There are also directory services which categorise feeds for you like technorati.com and syndic8.com.

We discussed AutoDiscovery which lets you know when a feed is present. If autoDiscovery is not present then rss/atom icons and links pointing to a feed are usually there somewhere.

General improvements

Content Syndication will continue to improve with ever improving aggregators, aggregators for mobile devices, Bittorrent being used for podcasts and videocasts to lighten the load on the hosting servers. More advanced polling for updates, using server ping, cloud technology (xml-rpc), trackbacks, delta polling.

Feeds will also become easier to find with the huge trend now in microformats, tagging/folksonomy (collaborative categorisation) of feeds, images, mp3s; the web will become easier to search. Content will be of better quality as the users will demand it and there will be less advertisements as the users will not have it.

General improvements

There has also been a huge surge in interest in social software like wikis, friendster, orkut etc. All of which support content syndication by default.

The number of blogging services is huge; blogger.com, msn spaces, wordpress.com, typepad, livejournal.com... all of which are huge proponents of Content Syndication through feeds and other means.

Content Syndication is bringing us closer to the ideal 'Semantic Web'.

The End